

# 基于TD3-PER的氢燃料电池混合动力汽车 能量管理策略研究\*

虞志浩<sup>1</sup> 赵又群<sup>1</sup> 潘陈兵<sup>1,2</sup> 何鲲鹏<sup>1,3</sup> 李丹阳<sup>1</sup>

(1.南京航空航天大学,南京 210016;2.安徽瑞氢动力科技有限公司,芜湖 241002;3.奇瑞新能源汽车股份有限公司,芜湖 241000)

**【摘要】**为优化氢燃料电池混合动力汽车的燃料经济性及辅助动力电池性能,提出了一种基于优先经验采样的双延迟深度确定性策略梯度(TD3-PER)能量管理策略。采用双延迟深度确定性策略梯度(TD3)算法,在防止训练过优估计的同时实现了更精准连续控制;同时结合优先经验采样(PER)算法,在获得更好优化性能的基础上加速了策略的训练。仿真结果表明:相较于深度确定性策略梯度(DDPG)算法,所提出的TD3-PER能量管理策略的百公里氢耗量降低了7.56%,平均功率波动降低了6.49%。

**关键词:**氢燃料电池混合动力汽车 优先经验采样 双延迟深度确定性策略梯度 连续控制  
**中图分类号:**U469.72 **文献标志码:**A **DOI:** 10.19620/j.cnki.1000-3703.20230350

## Research on Energy Management Strategy for Hydrogen Fuel Cell Hybrid Vehicles Based on TD3-PER

Yu Zhihao<sup>1</sup>, Zhao Youqun<sup>1</sup>, Pan Chenbing<sup>1,2</sup>, He Kunpeng<sup>1,3</sup>, Li Danyang<sup>1</sup>

(1. Nanjing University of Aeronautics and Astronautics, Nanjing 210016; 2. Anhui Ruiqing Power Technology Co., Ltd., Wuhu 241002; 3. Chery New Energy Vehicle Co., Ltd., Wuhu 241000)

**【Abstract】**To optimize the fuel economy and auxiliary power battery performance of hydrogen fuel cell hybrid vehicles, this article proposed an energy management strategy based on Twin Delayed Deep Deterministic policy gradient with Prioritized Experience Replay (TD3-PER). Adopting the Twin Delayed Deep Deterministic policy gradient (TD3) algorithm, it achieves more accurate continuous control while preventing training over estimation. By combining the Prioritized Experience Replay (PER) algorithm, the training of the strategy is accelerated while achieving better optimization performance. The simulation results show that compared with the Deep Deterministic Policy Gradient (DDPG) algorithm, the proposed TD3-PER energy management strategy reduces hydrogen consumption by 5.47% per 100 kilometers and average power fluctuation by 6.49%.

**Key words:** Hydrogen fuel cell hybrid vehicles, Prioritized Experience Replay (PER), Twin Delayed Deep Deterministic policy gradient (TD3), Continuous control

**【引用格式】**虞志浩, 赵又群, 潘陈兵, 等. 基于TD3-PER的氢燃料电池混合动力汽车能量管理策略研究[J]. 汽车技术, 2024(1): 13-19.

YU Z H, ZHAO Y Q, PAN C B, et al. Research on Energy Management Strategy for Hydrogen Fuel Cell Hybrid Vehicles Based on TD3-PER[J]. Automobile Technology, 2024(1): 13-19.

## 1 前言

氢燃料具有高能量密度和零污染等优点,以氢能源为动力的氢燃料电池汽车受到了越来越多的关注<sup>[1]</sup>。但

燃料电池存在瞬态响应慢和无法回收制动能量等缺点,通常需结合其他电源来满足车辆的行驶需求<sup>[1-4]</sup>。因此,有效的能量管理系统(Energy Management System, EMS)对不同电源能量的合理分配尤为重要。

\*基金项目:国家自然科学基金项目(52272397,11672127);中央高校基本科研业务费专项资金项目(NP2022408);高机动防暴车辆技术国家工程实验室开放基金项目(B20210017);南京航空航天大学科研与实践创新计划项目(xcxjh20220202)。

目前,能量管理策略主要分为基于规则的能量管理策略和基于优化的能量管理策略<sup>[5]</sup>。其中,基于规则的能量管理策略通常根据车辆行驶条件和车辆动力系统设计,简单实用<sup>[6]</sup>,但一般不具有最优性,较为依赖专家经验<sup>[7]</sup>。基于贝尔曼(Bellman)理论<sup>[8]</sup>的动态规划(Dynamic Programming, DP)方法作为最具代表性的全局优化算法,必须基于驾驶周期的先验信息进行优化求解,对不可预测和复杂驾驶条件的适应性较差<sup>[9]</sup>。

随着人工智能技术的快速发展,强化学习在混合动力汽车能量管理领域得到了广泛关注,因其具有自我适应、在线学习、试错和自我选择等特点而逐步成为氢燃料电池混合动力汽车能量管理的一种有效方法<sup>[10-11]</sup>。Q学习算法是强化学习领域的著名算法,Reddy等<sup>[12]</sup>采用Q学习算法来保持锂离子电池荷电状态(State Of Charge, SOC)的稳定性,提高了混合动力系统的效率。然而,Q学习算法在连续或多维状态空间中会大幅增加迭代计算次数。Wu等<sup>[13]</sup>进一步将深度Q学习算法应用于混合动力电动汽车的能量管理策略,与Q学习算法相比,该算法在模拟时间和收敛速度方面具有更好的性能,但其控制量需要降维和离散处理,进而使得控制精度下降。针对控制量离散的问题,Zhang等<sup>[14]</sup>提出了一种基于深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)的能量管理策略,进一步提高了燃料经济性,但该策略存在动作值过优估计的现象,进而导致训练不稳定等问题。

针对上述问题,同时为进一步提高车辆燃料经济性、改善电池SOC保持效果,本文提出基于优先经验采样的双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic policy gradient with Prioritized Experience Replay, TD3-PER)氢燃料电池混合动力汽车能量管理策略。采用双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic policy gradient, TD3)算法防止训练过优估计并更精准地进行连续控制,采用优先经验采样(Prioritized Experience Replay, PER)算法在获得更好优化性能的基础上加速策略的训练。

## 2 氢燃料电池汽车动力系统建模

本文所研究的燃料电池汽车动力系统型式为并联式,即燃料电池和锂离子电池分别通过DC/DC变换器和DC/AC变换器与功率总线相连,2种动力源均可单独为驱动电机供电。动力传动系统由燃料电池系统、锂离子电池、驱动电机、DC/DC变换器、DC/AC变换器、主减速器等组成,具体结构如图1所示,整车主要参数如表1所示。

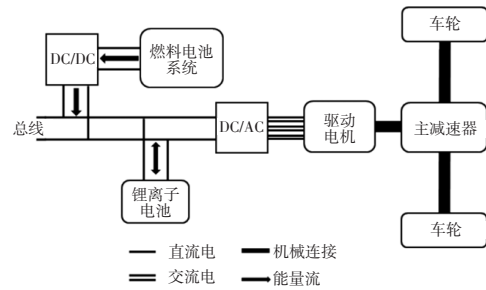


图1 氢燃料电池汽车动力系统构型

表1 整车主要参数

类别	参数	取值
整车	整车质量/kg	1 763
	滚动半径/m	0.301
	滚动阻力系数	0.013
	空气阻力系数	0.24
	迎风面积/m <sup>2</sup>	2.5
燃料电池系统	额定功率/kW	50
	峰值功率/kW	60
锂离子电池	容量/A·h	37
	总线额定电压/V	350
	最大电流/A	100
	持续电流/A	72
	荷电状态允许范围/%	5~95
驱动电机	峰值功率/kW	160
	最大转速/r·min <sup>-1</sup>	12 000

燃料电池系统通过单向DC/DC变换器后与总线相连,而不是直接与总线连接,避免了燃料电池系统电压因负载变化受到影响,对于延长燃料电池系统使用寿命及提高其效率具有良好的效果。通常,为降低功率相同时的电流损耗,燃料电池系统的输出电压会低于总线电压,所以DC/DC变换器一般为升压变换器。在功率波动相同时,低压侧电压波动更为平缓,因此有利于燃料电池系统工作效率的提高。而锂电池由于其电压波动允许范围较大,故可直接将其与总线并联。此系统构型兼顾了结构简单与可靠性,具有很强的可控性,已得到广泛应用<sup>[14]</sup>,满足本文要求。

### 2.1 氢燃料电池混合动力汽车的需求功率

在工况已知时,总需求功率 $P_r$ 可由车辆的动力学平衡方程计算得出:

$$\begin{cases} P_r = Mgf v \cos \alpha + Mgv \sin \alpha + mav\delta + 0.5\rho A_f C_d v^2 \\ P_m = P_{req} / \eta_{motor} \\ P_m = P_{fc} \cdot \eta_{fc} + P_{batt} \end{cases} \quad (1)$$

式中, $M$ 为整车总质量; $g$ 、 $f$ 、 $v$ 分别为重力加速度、滚动阻力系数和车速; $\alpha=0$ 为道路坡度; $\rho$ 为空气密度; $A_f$ 为迎

风面积;  $C_D$  为空气阻力系数;  $\delta=1$  为质量系数;  $a$  为车辆加速度;  $P_{req}$ 、 $P_m$ 、 $P_{fc}$ 、 $P_{batt}$  分别为车辆的驱动功率、电机需求功率、燃料电池和锂离子电池的输出功率;  $\eta_{motor}$ 、 $\eta_{fc}$  分别为电机的效率和燃料电池的效率。

## 2.2 氢燃料电池系统模型

氢燃料电池组是燃料电池混合动力汽车的主要动力源,它通过电化学反应将化学能转化为电能。受一些部件的启动特性和物理特性影响,氢燃料电池组在低功率范围内的效率较低,而在高功率范围内效率仍会降低。氢耗  $C_{fc}$  和燃料电池效率  $\eta_{fc}$  的计算方程为:

$$\begin{cases} C_{fc} = \frac{N_{cell} \cdot M(H_2)}{2F} \cdot I_{stack} \cdot \lambda \\ \eta_{fc} = \frac{P_{fc}}{\dot{m} \cdot L_{H_2}} \end{cases} \quad (2)$$

式中,  $N_{cell}$  为氢燃料电池组中电池的数量;  $M(H_2)=2$  g/mol 为氢气的摩尔质量;  $F$  为法拉第常数;  $I_{stack}$  为燃料电池电流;  $\lambda=1.05$  为氢过量比;  $\dot{m}$  为氢的质量变化量;  $L_{H_2}=120$  mJ/kg 为氢的热值。

能量管理策略对燃料电池组的寿命具有重要影响<sup>[15]</sup>,包括催化剂损失、燃料渗透和质子交换膜降解等方面。其中,燃料电池组降解率直接关系到电压的衰减,Pu等<sup>[16]</sup>提出了单个燃料电池的电压退化模型。燃料电池退化经验公式为:

$$D_{fc} = k_1 t_1 + k_2 t_2 + k_3 \left| \frac{P_{t+1} - P_t}{P_{high} - P_{low}} \right| \quad (3)$$

式中,  $D_{fc}$  为燃料电池退化率;  $t_1$ 、 $t_2$  分别为高负荷、低负荷运行时间;  $k_1$ 、 $k_2$ 、 $k_3$  分别为高负荷、低负荷和负荷变化引起的退化率;  $P_{high}$ 、 $P_{low}$  分别为燃料电池组输出功率的最大值和最小值;  $P_t$ 、 $P_{t+1}$  分别为燃料电池组的即时输出功率和下一时刻的输出功率。

## 2.3 锂离子电池模型

锂离子电池具有能量密度高、自放电率低和寿命长等优点,是燃料电池混合动力汽车的理想辅助电源。锂离子电池采用  $R_{int}$  内阻等效电路模型:

$$\begin{cases} P_{batt}(t) = V_{oc}(t) - R_0 \cdot I^2(t) \\ I(t) = \frac{V_{oc}(t) - \sqrt{4R_0 \cdot P_{batt}(t)}}{2R_0} \\ S_{soc}(t) = \frac{Q_0 - \int I(t) dt}{Q} \end{cases} \quad (4)$$

式中,  $P_{batt}(t)$ 、 $V_{oc}(t)$ 、 $R_0$ 、 $I(t)$  分别为锂离子电池的功率、开路电压、内阻和电流;  $S_{soc}(t)$  为电池的荷电状态;  $Q_0$ 、 $Q$  分别为锂离子电池的初始容量和额定容量。

为了更好地反映车辆在行驶过程中的燃料消耗量,锂离子电池当量氢消耗量  $C_{batt}$  可表示为:

2024年 第1期

$$C_{batt} = \begin{cases} \frac{P_{batt} \eta_{DC} \eta_1 M(H_2)}{\Delta H}, P_{batt} < 0 \\ \frac{P_{batt} M(H_2)}{\eta_{DC} \eta_1 \Delta H}, P_{batt} \geq 0 \end{cases} \quad (5)$$

式中,  $\eta_{DC}=0.9$  为DC/DC变换器效率;  $\Delta H=240$  kJ/mol 为氢的低热值;  $\eta_1=0.9$  为锂离子电池的放电、充电效率。

## 3 基于TD3-PER算法的能量管理策略设计

### 3.1 TD3-PER算法

TD3是由Scott Fujimoto等人在DDPG算法基础上改进得到的一种用于解决连续控制问题的在线(On-Line)异策(Off-Policy)式深度强化学习算法,深度强化学习框架如图2所示。

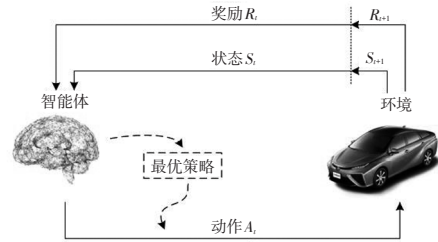


图2 深度强化学习框架

TD3算法是演员-评论家(Actor-Critic, AC)框架下的一种确定性深度强化学习算法,其中策略网络(Actor)以车辆的状态作为输入并根据网络参数输出控制动作,评价网络(Critic)用于评价策略网络所执行动作的优劣。TD3算法能同时处理连续动作空间和策略值函数过优估计的问题,图3所示为TD3算法框架。

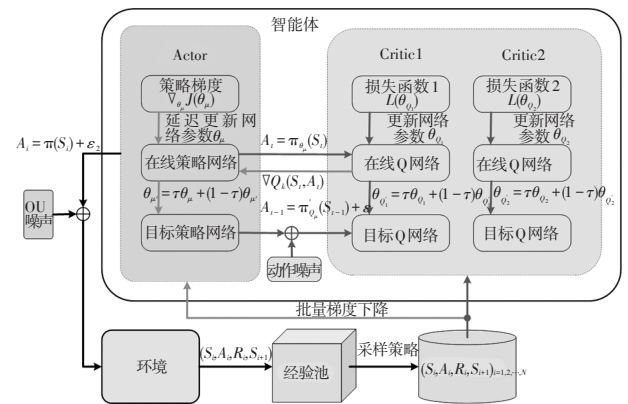


图3 TD3算法框架

TD3算法中包括6个网络,分别是Actor网络  $\mu(\cdot|\theta_\mu)$ 、Critic1网络  $Q_1(\cdot|\theta_{Q_1})$ 、Critic2网络  $Q_2(\cdot|\theta_{Q_2})$ 、Actor目标网络  $\mu'(\cdot|\theta_{\mu'})$ 、Critic1目标网络  $Q'_1(\cdot|\theta_{Q'_1})$ 、Critic2目标网络  $Q'_2(\cdot|\theta_{Q'_2})$ 。TD3算法建立了2个独立的Critic网络,在计算目标值时取二者中的较小值,从而抑制网络过估计问题:

$$y_t = R_t + \gamma \min_{k=1,2} Q'_k(S_{t+1}, \tilde{A} | \theta_{Q'_k}) \quad (6)$$

式中,  $R_t$  为  $t$  时刻即时奖励;  $\gamma$  为折扣因子;  $y_t$  为累计奖励;  $S_{t+1}$  为  $(t+1)$  时刻状态;  $\tilde{A}$  为加入噪声后的目标动作。

确定性策略存在一个问题, 即会过度拟合以减小价值估计中的峰值。当更新 Critic 网络时, 使用确定性策略的学习目标极易受到函数逼近误差的影响, 从而导致目标估计的方差偏大, 估计值不准确。这种诱导方差可以通过正则化来减小, 在实际操作时为向目标动作中添加服从截断正态分布的噪声  $\varepsilon$ :

$$\begin{cases} \tilde{A} \leftarrow \mu'(S_{t+1} | \theta_{\mu'}) + \varepsilon \\ \varepsilon \sim \text{clip}(N(0, \sigma), -c, c), c > 0 \end{cases} \quad (7)$$

式中,  $c$  为截断值;  $\sigma$  为标准差。

将式(6)代入贝尔曼方程, 计算损失函数:

$$L(\theta_{Q'_k}) = \frac{1}{M} \sum_{i=1}^M [y_i + Q_k(S_i, A_i | \theta_{Q'_k}) - Q_k(S_i, A_i)]^2 \quad (k=1, 2) \quad (8)$$

式中,  $M$  为 Critic 网络更新周期;  $\theta_{Q'_k}$  为 Critic 网络的权重参数;  $A_i$  为  $t$  时刻动作。

目标函数为:

$$\begin{cases} J(\theta_{Q'_k}) = \min_{\theta_{Q'_k}} E_{\pi} \left[ \frac{1}{2} (Q_k(S_i, A_i) - y_i)^2 \right] \\ J(\theta_{\mu'}) = \max_{\theta_{\mu'}} E_{\pi} [Q_k(S_i, A_i)] \end{cases} \quad (k=1, 2) \quad (9)$$

式中,  $\theta_{\mu'}$  为 Actor 网络的权重参数;  $E_{\pi}$  为执行策略  $\pi$  得到的期望回报。

Critic 网络的权重参数  $\theta_{Q'_k}$  通过梯度下降法最小化损失函数  $L(\theta_{Q'_k})$  来更新:

$$\begin{aligned} \nabla_{\theta_{Q'_k}} J(\theta_{Q'_k}) &= \nabla_{\theta_{Q'_k}} E \left[ \frac{1}{2} (Q_k(S_i, A_i) - (R_i + \gamma Q'_k(S_{i+1}, A_{i+1})))^2 \right] \\ &= \frac{1}{M} \sum_{i=1}^M \left[ (R_i + \gamma Q'_k(S_{i+1}, A_{i+1}) - Q_k(S_i, A_i)) \nabla_{\theta_{Q'_k}} Q_k(S_i, A_i) \right] \end{aligned} \quad (k=1, 2) \quad (10)$$

式中,  $\nabla$  为权重参数的梯度;  $E$  为期望回报。

Actor 网络参数  $\theta_{\mu'}$  的在线更新通过梯度上升法使  $Q$  值关于  $\theta_{\mu'}$  迭代增加:

$$\nabla_{\theta_{\mu'}} J(\theta_{\mu'}) = \frac{1}{N} \sum_i \left[ \nabla_{A_i} Q_k(S_i, A_i) \cdot \nabla_{\theta_{\mu'}} \pi(S_i) \right] \quad (k=1, 2) \quad (11)$$

式中,  $N$  为样本数量;  $\pi$  为执行策略。

由于 Actor 网络通过最大化累积期望回报来更新, 它需要利用 Critic 网络进行评估。如果 Critic 网络非常不稳定, 则 Actor 网络也会出现振荡。因此, 可使 Critic 网络的更新频率高于 Actor 网络的更新频率, 即等待 Critic 网络更加稳定后再帮助 Actor 网络更新, 通过延迟更新 (Delayed Update) 的方式保证 Actor 网络的训练更加稳定。

目标网络的更新采用软更新方式, 引入学习率  $\tau$ , 将旧的目标网络参数和新的对应网络参数进行加权平均,

然后赋值给目标网络:

$$\theta_{Q'_k} = \tau \theta_{Q'_k} + (1 - \tau) \theta_{Q'_k} \quad (k=1, 2) \quad (12)$$

$$\theta_{\mu'} = \tau \theta_{\mu'} + (1 - \tau) \theta_{\mu'} \quad (13)$$

为使网络训练速度更快、效果更好, 本文利用 PER 算法, 并结合训练经验赋予一定的权重, 如在交互过程中对表现较差的片段赋予更高权重, 使这些片段被网络重新学习的概率更高, 从而大幅提高模型的学习效率。本文通过时序差分误差 (Temporal Difference error, TD-error) 的绝对值  $|\sigma|$  对片段的重要性进行表征,  $|\sigma|$  越大, 表示该片段的重要程度越高:

$$|\sigma| = |R_{t+1} + \gamma Q'(S_{t+1}, A_{t+1}) - Q(S_t, A_t)| \quad (14)$$

式中,  $Q'(S_{t+1}, A_{t+1})$  为根据  $(t+1)$  时刻的状态值  $S_{t+1}$ 、动作值  $A_{t+1}$  和网络参数  $\theta_{Q'}$  得到的目标  $Q$  网络的  $Q$  值;  $Q(S_t, A_t)$  为根据  $t$  时刻的状态值  $S_t$ 、动作值  $A_t$  和网络参数  $\theta_Q$  得到的目标  $Q$  网络的  $Q$  值。

将经验样本依据  $|\sigma|$  排序, 得到样本  $m$  的序列  $r(m)$ , 进而获得优先级指标  $D_m$ :

$$D_m = 1/r(m) > 0 \quad (15)$$

然而, 直接采用上述采样方式会造成贪婪抽样, 使得初始  $|\sigma|$  较小时在较长时间段内不被抽样, 以及  $|\sigma|$  较大时被高频重复抽样等问题, 继而影响样本多样性。针对此问题, 本文采用贪婪抽样与均匀采样相结合的方式, 设定经验池中每个样本  $m$  的采样概率  $P(m)$  为:

$$P(m) = \frac{D_m^\alpha}{\sum_{k=1}^V D_k^\alpha} \quad (16)$$

式中,  $D_m$  为第  $m$  个样本的优先级指标;  $\alpha$  为超参数, 均匀抽样时  $\alpha=0$ , 贪婪抽样时  $\alpha=1$ , 2 种采样结合时  $\alpha \in (0, 1)$ ;  $V$  为样本池大小。

另一方面, 使用 PER 算法后, 可能导致样本分布改变, 进而使模型收敛到不同的值。针对此问题, 本文引入重要性采样, 使每个样本以不同概率被选取, 并在提升训练速度的基础上, 保障对梯度下降的相同影响, 从而解决模型收敛的问题。重要性采样中样本  $m$  的权重  $W_m$  为:

$$W_m = \left[ \frac{1}{N} \cdot \frac{1}{P(m)} \right]^\beta \quad (17)$$

式中,  $N$  为经验池中的样本数量;  $\beta \in [0, 1]$  为超参数, 用来表征 PER 算法对收敛结果的影响程度,  $\beta=1$  时 PER 算法对收敛结果无任何影响。

### 3.2 基于 TD3-PER 算法的能量管理策略

本文选取车速  $v$ 、车辆加速度  $a$  和电池荷电状态  $S_{\text{soc}}$  作为状态变量, 即  $S_t = \{v, a, S_{\text{soc}}\}$ , 动作变量  $A_t$  为燃料电池系统输出功率。文献[17]揭示了输出功率波动会极大

地影响燃料电池的寿命,为了最大限度地减少氢消耗,防止锂离子电池的过度充电和过度放电,并延长燃料电池的寿命,本文将奖励函数定义为:

$$R_t = -\left\{ \chi \dot{m}(t) + \beta [S_{\text{SOCref}} - S_{\text{SOC}}(t)]^2 + \omega [P_t - P_{t-1}]^2 \right\} \quad (18)$$

式中, $\chi$ 、 $\beta$ 、 $\omega$ 分别为氢耗量、锂离子SOC维持、燃料电池系统输出功率变化的权重; $S_{\text{SOCref}}$ 为SOC的参考值(本文取0.65)。

氢燃料电池汽车为能量管理策略中的环境,智能体根据汽车状态 $S_t$ 和智能体中的策略 $\pi^*$ 在每一步为车辆选择一个动作 $A_t$ ,车辆反馈即时奖励 $R_t$ 和下一刻状态 $S_{t+1}$ 。将 $(S_t, A_t, R_t, S_{t+1})$ 存入经验池(Replay Buffer),形成历史数据,经验池的使用可以有效消除相邻状态间的相关性。为提高网络的收敛速度和训练效果,采用优先经验采样的方式从经验池中抽取历史数据在智能体的网络中进行训练,同时,智能体在与环境的交互中不断调整网络权重得到最优策略 $\pi^*$ 。

#### 4 仿真与验证

为了验证TD3-PER算法在燃料电池混合动力车辆的燃料消耗、锂离子SOC维持和减少输出功率波动方面的性能,在新欧洲驾驶循环(New European Driving Cycle, NEDC)工况下应用TD3-PER算法进行仿真分析,并将其与基于DDPG算法的能量管理策略进行对比。TD3-PER算法的主要超参数如表2所示。

表2 TD3-PER算法主要超参数

参数	取值
折扣因子	0.9
经验池大小	$10^4$
Actor网络学习率	0.001
Critic网络学习率	0.001
样本学习数量/个	64
训练步数/步	2 000
软更新参数	0.001
延迟更新参数	2

在深度强化学习的迭代过程中,将奖励均值作为收敛标准和性能评价指标。训练初期,智能体会不断探索未知的环境以获得更好的训练效果;随着训练的进行,智能体会更倾向于选择奖励更高的动作,使得奖励均值逐渐收敛。2种算法的奖励均值如图4所示,基于TD3-PER算法与基于DDPG算法的能量管理策略分别第360步和第620步左右收敛,并且TD3-PER算法的奖励均值波动更小,训练效果更为稳定。可以看出,与DDPG

算法相比,TD3-PER算法具有更快的收敛速度,且训练效果更好。因此,TD3-PER算法可以通过较少的迭代,快速优化燃料消耗量并缓解燃料电池堆的功率波动。

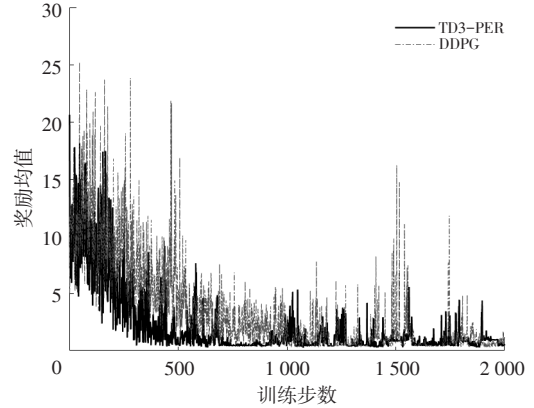


图4 TD3-PER算法和DDPG算法的奖励均值

NEDC工况车速如图5所示,2种策略燃料电池系统输出功率、锂离子电池输出功率和锂离子电池SOC的比较结果如图6所示,不同策略下的燃料经济性与平均功率波动差异对比结果如表3所示。

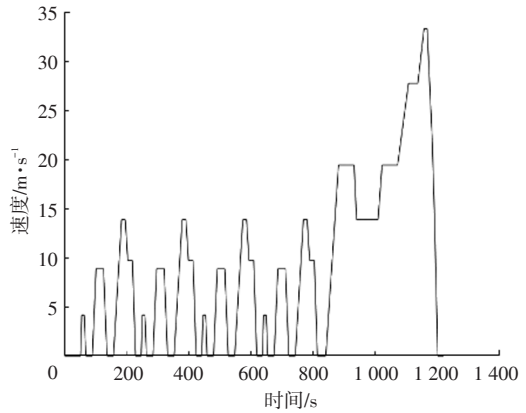
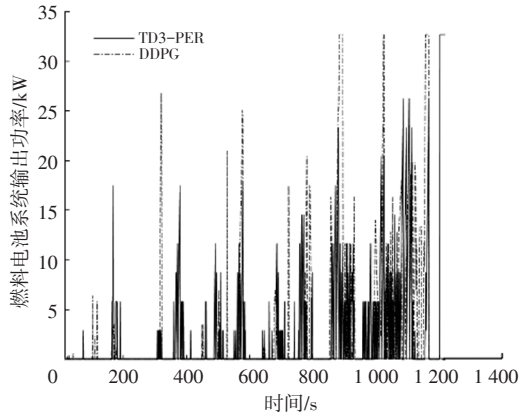


图5 NEDC工况车速

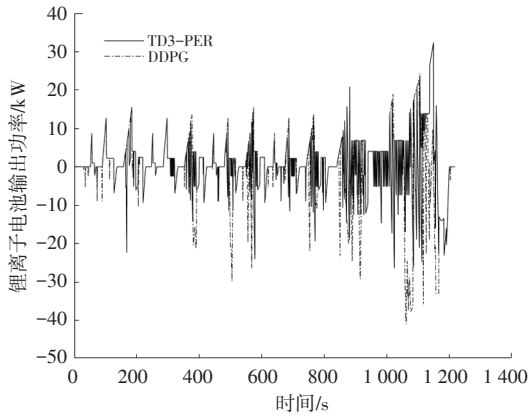
由表3可知,将整个循环工况氢耗总量等比例换算为百公里氢耗量后,TD3-PER策略相较于DDPG策略百公里氢耗量降低了7.56%,有效提高了燃料经济性。由于在设置奖励函数时将燃料电池功率变化也考虑在内,在对策略进行优化时,TD3-PER算法能在保证动力性的同时限制燃料电池功率的剧烈变化,从图6a中可以看出,TD3-PER策略的燃料电池系统功率输出相较于DDPG策略更为平缓,同时从表3的对比结果可以看出,TD3-PER策略平均功率波动相比DDPG策略降低了6.49%,这表明基于TD3-PER算法的能量管理策略能够有效降低动态加载工况对燃料电池使用寿命的影响,从而提高车辆耐久性。

由图6b可知,锂离子电池作为二次电源,在高车辆负载和需求功率瞬时变化时可快速响应,实现缓解功率波动并延长燃料电池组寿命的作用。本文在设置奖励

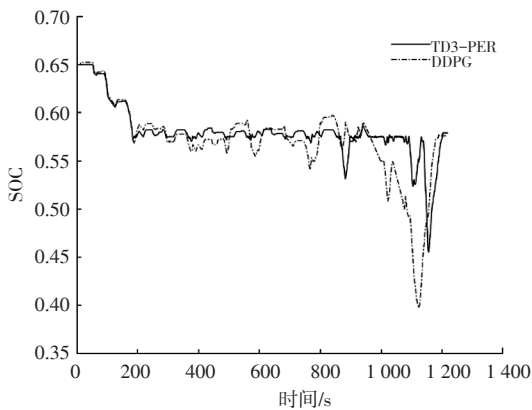
函数时还考虑了锂离子电池SOC的维持因子,以保证较为稳定的锂离子电池工作状态。由图6c可以看出,TD3-PER策略的SOC曲线在整个工况区间内波动相较于DDPG策略更为平缓,有效抑制了锂离子电池过充或过放,能够提高锂离子电池寿命。



(a)燃料电池系统输出功率



(b)锂离子电池输出功率



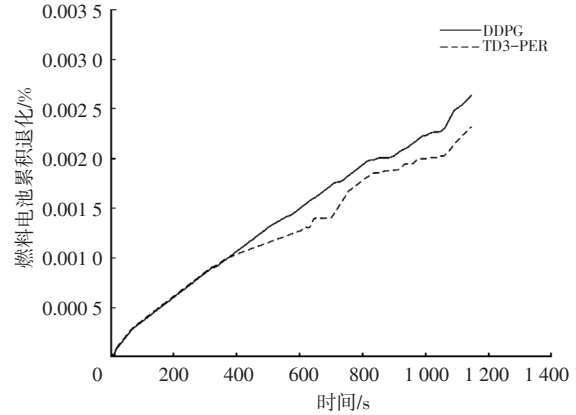
(c)锂离子电池SOC

图6 TD3-PER算法和DDPG算法下燃料电池系统输出功率、锂离子电池输出功率及锂离子电池SOC

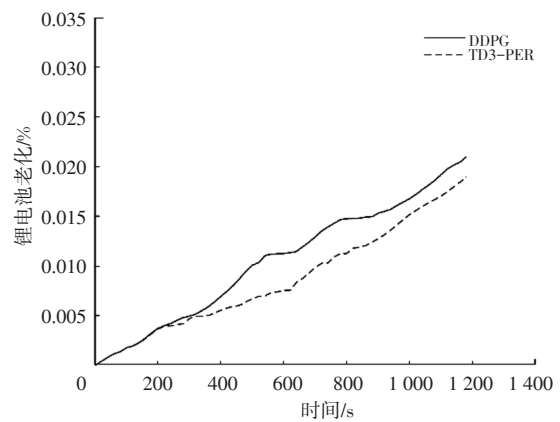
表3 不同策略下的燃料经济性与平均功率波动对比

算法	SOC终值	百公里氢耗量/kg	平均功率波动/kW
TD3-PER	0.578	883.861	1.715
DDPG	0.576	956.163	1.834

为了验证不同算法在NEDC工况下的燃料电池寿命和锂电池寿命,本文分别以燃料电池的累积退化与锂电池的老化作为燃料电池寿命与锂电池寿命的评估指标,对TD3-PER算法和DDPG算法下的寿命进行了比较,结果如图7所示。



(a)燃料电池累积退化



(b)锂电池老化

图7 TD3-PER算法和DDPG算法下燃料电池和锂电池寿命

在NEDC工况循环结束时:基于TD3-PER算法和DDPG算法的燃料电池累计退化分别为0.00229%和0.00262%,基于TD3-PER算法的燃料电池累计退化比DDPG算法低12.6%;基于TD3-PER算法和DDPG算法的锂电池老化分别为0.0189%和0.0215%,基于TD3-PER算法的锂电池老化比DDPG算法低12.1%。由此可以看出,相比于DDPG算法,TD3-PER算法在提高燃料电池寿命及锂电池寿命方面表现更好。

## 5 结束语

本文提出了一种基于TD3-PER算法的氢燃料电池混合动力汽车能量管理策略,并在NEDC工况下对基于TD3-PER算法的能量管理策略在燃料电池混合动力车辆的燃料经济性、锂电池SOC维持和减少输出功率波动方面的性能,得到以下结论:

a. 与DDPG算法相比,TD3-PER算法能在更短的训练步数内完成收敛,且训练效果更为稳定,在一定程度上提高了优化效率及效果。

b. TD3-PER策略相较于DDPG策略百公里氢耗量降低了7.56%,有效提高了燃料经济性。

c. TD3-PER策略的SOC在整个工况区间内波动相较于DDPG策略更为平缓,有效抑制了锂离子电池过充或过放,提高了锂离子电池寿命。同时,TD3-PER策略平均功率波动相比DDPG策略降低了6.49%,表明基于TD3-PER算法的能量管理策略能够有效降低动态加载工况对于燃料电池使用寿命的影响,从而提高车辆耐久性。

为使TD3-PER算法更好地适应复杂的行驶工况,下一步工作将对在线更新网络参数开展研究,以进一步提高TD3-PER算法对复杂工况的适应性和整车燃料经济性。

### 参 考 文 献

- [1] WU J D, WEI Z B, LI W H, et al. Battery Thermal- and Health-Constrained Energy Management for Hybrid Electric Bus Based on Soft Actor-Critic DRL Algorithm[J]. IEEE Transactions on Industrial Informatics, 2021, 17(6): 3751-3761.
- [2] HAN X, LI F Q, ZHANG T, et al. Economic Energy Management Strategy Design and Simulation for a Dual-Stack Fuel Cell Electric Vehicle[J]. International Journal of Hydrogen Energy, 2017, 42(16): 11584-11595.
- [3] ZUO W, LI J, ZHANG Y T, et al. Effects of Multi-Factors on Comprehensive Performance of a Hydrogen-Fueled Micro-Cylindrical Combustor by Combining Grey Relational Analysis and Analysis of Variance[J]. Energy, 2020, 199.
- [4] WANG Y J, SUN Z D, CHEN Z H. Energy Management Strategy for Battery/Supercapacitor/Fuel Cell Hybrid Source Vehicles Based on Finite State Machine[J]. Applied Energy, 2019, 254.
- [5] LÜ X Q, QU Y, WANG Y D, et al. A Comprehensive Review on Hybrid Power System for PEMFC-HEV: Issues and Strategies[J]. Energy Conversion and Management, 2018, 171: 1273-1291.
- [6] ZHANG J Q, JIAO X H, YANG C. A Double-Deep Q-Network-Based Energy Management Strategy for Hybrid Electric Vehicles under Variable Driving Cycles[J]. Energy Technology, 2021, 9(2).
- [7] CHEN Z, HU H J, WU Y T, et al. Stochastic Model Predictive Control for Energy Management of Power-Split Plug-in Hybrid Electric Vehicles Based on Reinforcement Learning [J]. Energy, 2020, 211.
- [8] LIU Y G, HUANG Z Z, LI J, et al. Cooperative Optimization of Velocity Planning and Energy Management for Connected Plug-in Hybrid Electric Vehicles[J]. Applied Mathematical Modelling, 2021, 95: 715-733.
- [9] TAN H C, ZHANG H L, PENG J K, et al. Energy Management of Hybrid Electric Bus Based on Deep Reinforcement Learning in Continuous State and Action Space[J]. Energy Conversion and Management, 2019, 195: 548-560.
- [10] 周维, 张承宁, 李军求. 增程式电动车BL和CD-CS型最优能量管理策略比较研究[J]. 汽车工程, 2016, 38(12): 1407-1414.
- ZHOU W, ZHANG C N, LI J Q. A Study on Comparison between BL and CD-CS Optimal Energy Management Strategies for Extended Range Electric Vehicles[J]. Automotive Engineering, 2016, 38(12): 1407-1414.
- [11] 王哲, 谢怡, 臧鹏飞, 等. 基于极小值原理的燃料电池客车能量管理策略[J]. 吉林大学学报(工学版), 2020(1): 36-43.
- WANG Z, XIE Y, ZANG P F, et al. Energy Management Strategy for Fuel Cell Buses Based on Minimum Principle [J]. Journal of Jilin University (Engineering and Technology Edition), 2020(1): 36-43.
- [12] REDDY N P, PASDELOUP D, ZADEH M K, et al. An Intelligent Power and Energy Management System for Fuel Cell/Battery Hybrid Electric Vehicle Using Reinforcement Learning[C]// 2019 IEEE Transportation Electrification Conference and Expo (ITEC). Detroit, MI, USA: IEEE, 2019.
- [13] WU J D, HE H W, PENG J K, et al. Continuous Reinforcement Learning of Energy Management with Deep Q Network for a Power Split Hybrid Electric Bus[J]. Applied Energy, 2018, 222: 799-811.
- [14] ZHANG B, WU J L, ZOU Y, et al. Reinforcement Learning Energy Management for Hybrid Electric Tracked Vehicle with Deep Deterministic Policy Gradient[C]// Proceedings of China SAE Congress 2020: Selected Papers. Shanghai: Springer, 2022.
- [15] SONG K, CHEN H, WEN P M, et al. A Comprehensive Evaluation Framework to Evaluate Energy Management Strategies of Fuel Cell Electric Vehicles[J]. Electrochimica Acta, 2018, 292: 960-973.
- [16] PU Y C, LI Q, ZOU X L, et al. Optimal Sizing for an Integrated Energy System Considering Degradation and Seasonal Hydrogen Storage[J]. Applied Energy, 2021, 302(1).
- [17] INCI M, BÜYÜK M, DEMIR M H, et al. A Review and Research on Fuel Cell Electric Vehicles: Topologies, Power Electronic Converters, Energy Management Methods, Technical Challenges, Marketing and Future Aspects[J]. Renewable and Sustainable Energy Reviews, 2021, 137.

(责任编辑 斛 畔)

修改稿收到日期为2023年7月16日。